Amendments to the Claims:

Listing of Claims:

1. (Canceled).
2. (Canceled).
3. (Canceled).
4. (Canceled).
5. (Canceled).
6. (Canceled).
7. (Canceled).
8. (Canceled).
9. (Canceled).
10. (Canceled).
11. (Canceled).
12. (Canceled).
13. (Canceled).
14. (Canceled).
15. (Original) In a system comprising a proxy gateway connected by a first network to a plurality of mobile users and by a second network to at least one Web server, said proxy gateway comprising a cache for storing pull content received from said at least one Web server of a pull service, a method comprising the steps of:

storing data that is indicative of a request for said pull content from at least one of said plurality of mobile users and data indicative of interactions between said cache and said Web server;

determining access probability of access to said pull content from said stored data;

determining an average hit rate for said pull content from said stored data;

determining said average response delay for said pull content from said stored data;

determining average wired network access latency for said pull content from said access probability, said average hit rate and said average response delay;

storing said pull content in said cache based on said determined average wired network access latency when there is no said pull content in said cache or said pull content has expired,

wherein said pull content having a greater average wired network access latency is prioritized for being stored in said cache.

16. (Original) The method of claim 15 wherein said pull content is a plurality of n documents, n=1, 2 . . . N, wherein N is the total number of documents, and said stored data comprises:

an average rate of access to document n, R_n; a size of said document n, s_n; an average time delay imposed by said second network, ΔT_n; and an update cycle of said document n, μ_n.

17. (Original) The method of claim 16 wherein said access probability is determined by:

$\gamma_n = R_n / R$

wherein R is the total rate of access traffic on said second network.

18. (Original) The method of claim 17 wherein said average hit rate for document n, h_n is determined by:

$$h_n = 1 - \frac{g_n}{R_n \mu_n}, \quad n=1, 2, \ldots, N,$$

in which:

g_n is the probability that there is at least one request to document n during a given update cycle, μ_n, given by:

$g_n = 1 - e^{-R_n \mu_n}, \quad n = 1, 2, \ldots, N,$ (3)

and

$R_n \mu_n$ is the expected number of accesses to document n in an update cycle of document n.

19. (Original) The method of claim 18 wherein average wired-network-access latency when there is no said pull content in said cache or said pull content has expired is determined from

$\eta_n = \gamma_n (1-h_n) \Delta T_n, \quad n = 1, \ldots, N,$

20. (Original) The method of claim 19 wherein said pull content is prioritized by the steps of:

sorting said plurality of N documents in descending order with the document having the greatest average wired network access latency when there is no said pull content in said cache or said pull content has expired labeled as η_1, and the document having the least average wired-network-access latency when there is no said Web content in said cache or said Web content has expired labeled as η_N; and

determining a number of documents to be stored in said cache, r, by considering at least one constraint selected from the group consisting of spare cache capacity, spare transmission bandwidth on said second network and desired hit probability.

21. (Original) The method of claim 20 wherein said constraint of said spare cache capacity, ΔC, is given by:

$$\sum_{n=1}^{r} s_n (1-h_n) \leq \Delta C,$$

wherein $\Delta C \approx C - \sum_{n=1}^{N} s_n h_n$, C is given capacity of the cache, s_n is the size of the document n and h_n is the average hit rate for document n.

22. (Original) The method of claim 20 wherein said constraint of said spare transmission bandwidth, ΔB, is given by:

$$\sum_{n=1}^{r} (1-g_n) \frac{s_n}{\mu_n} \leq \Delta B,$$

wherein $\Delta B \approx B - \sum_{n=1}^{N} g_n \frac{s_n}{\mu_n}$, B is given bandwidth, g_n is the probability that there is at least one request to document n during a given update cycle μ_n and s_n is the size of the document.

23. (Original) The method of claim 20 wherein said constraint of said desired minimum hit probability, ΔH, is given by:

$$\sum_{n=1}^{r} \gamma_n (1-h_n) \geq \Delta H$$

wherein $\Delta H \approx H - \sum_{n=1}^{N} \gamma_n h_n$ H is given hit probability, γ_n is an access to document n and h_n is an average hit rate for document n.

24. (Original) The method of claim 16 further comprising the step of:

updating said stored pull content in said cache based on said update cycle of document n, μ_n.

25. (Original) In a system comprising a proxy gateway connected by a first network to a plurality of mobile users and by a second network to at least one Web server, a method comprising the steps of:

measuring each of said mobile users current geo-location position and behavior;

computing a first probability that said measured current geo-location position and behavior is an actual position and behavior of each of said mobile users;

determining a state sequence estimation variable for each of said mobile users by iteration over time from a second probability that each of said mobile users transit in a geo-location and behavior sequence;

determining a current state for each of said mobile users from said state sequence estimation; and

pushing push content related to said current state to each of said mobile users.

26. (Original) The method of claim 25 wherein said first probability is given by

$P_r \{Y_t \mid X\}$

wherein Y_t is said measured current geo-location position and behavior and X is said actual position and behavior.

27. (Original) The method of claim 26 wherein said state sequence estimation variable is determined by

$$\alpha_t(m) = \sum_{m'=0}^{M-1} \alpha_{t-1}(m')\, p_{m'm} \sum_x \Pr\{x \mid m\} \Pr\{Y_t \mid x\}$$

wherein $p_{m'm}$ is the state transition probability of one of said plurality of mobile users, $\Pr\{x \mid m\}$ is the probability that said one of said plurality of mobile users locates at position and behavior as x when it is in state m at time t, and $\Pr\{Y_t \mid x\}$ is the probability that said measured geo-location is Y_t when said one of said plurality of mobile users position and behavior is x at time t.

28. (Original) The method of claim 27 wherein said current state is determined by

$z = \arg \max_m \{\alpha_t(m) \mid m = 1,2, \ldots, M-1\}$.

29. (Original) The method of claim 28 wherein said proxy gateway comprising a cache for storing push content received from said at least one Web server and said push content is stored in said cache based on said current state.